Scientometrics
https://doi.org/10.1007/s11192-018-2973-6

Scholarly impact assessment: a survey of citation weighting solutions

Liwei Cai[1] · Jiahao Tian[1] · Jiaying Liu[1] · Xiaomei Bai[2] · Ivan Lee[3] · Xiangjie Kong[1] · Feng Xia[1]

Received: 18 June 2018

Abstract
Scholarly impact assessment has always been a hot issue. It has played an important role in evaluating researchers, scientific papers, scientific teams, and institutions within science of science. Scholarly impact assessment is also used to address fundamental issues, such as reward evaluation, funding allocation, promotion and recruitment decision. Scholars generally agree that it is more reasonable to use weighted citations to assess the scholarly impact. Although a great number of researchers use weighted citations to access the scholarly impact, there is a lack of a systematic summary of citation weighting methods. To fill the gap, this paper summarizes the existing classical indicators and weighting methods used in measuring scholarly impact from the perspectives of articles, authors and journals. We also summarize the focus of the indicators involved in this paper and the weighting factors that involved in the weighting methods. Finally, we discuss the open issues to try to discover the hidden trends of citation weighting. Through this paper, we can not only have a clearer understanding of the weighting methods in the scholarly impact assessment, but also think more deeply about the weighting factors to be explored.

Keywords Weighted citations · Scholarly impact assessment · Weighting factors

Introduction

Scholarly impact refers to publishing new academic theories, methods or techniques, or developing, modifying, reforming and perfecting the original achievements, so as to influence the theory and practice in related fields (Van et al. 2000; Kong et al. 2018). Scholarly impact includes the impact of articles, authors, journals, institutions, etc. The scholarly impact assessment plays an important role in quantifying the

✉ Xiangjie Kong
xjkong@ieee.org

1 Key Laboratory for Ubiquitous Network and Service Software of Liaoning Province, School of Software, Dalian University of Technology, Dalian, China

2 Computing Center, Anshan Normal University, Anshan, China

3 School of Information Technology and Mathematical Sciences, University of South Australia, Adelaide, Australia

Published online: 10 December 2018

scientific contributions of researchers, teams, institutions, and countries. It is also used to address the following fundamental problems, such as the reward evaluation, funding allocation, promotion and recruitment decision (Bai et al. 2017b).

Bollen et al. (2009) point out that the science is a *gift economy*. The value of science is defined as the contribution of one's mind to knowledge or to the ideas of others. Authors use citations to acknowledge publications that influenced their work. So the scholarly impact can be measured in terms of citations that a publication receives. Traditionally, the importance of a research article is measured by its citations. This approach has drawn criticism: If we consider citations as scholarly votes (Davis 2008), the question is, should all votes be treated equally? Citations are unweighted in traditional citation analysis. Treating all citations equally ignores the wide variety of functions that citations perform (Zhu et al. 2015). For example, an article cited by a more influential journal is more likely to have a higher *real* impact. For a researcher, his/her research is more valuable if his/her articles are cited by more authoritative researchers. It is similar for journals. Furthermore, there are other issues with raw citations. For example, as citations of an article or a journal span over time, should old citations be treated the same as new ones? Moreover, authors tend to cite their own work, and in general these self-citations are less valuable. With the development of scientific collaboration, an article may be written by multiple authors, do all authors make equal contributions to the article? It has also been found that citation patterns significantly vary across different research fields. So it's not reasonable to directly compare the citations of journals from different fields(Vaccario et al. 2017). Last, but not least, that negative citations (Cavalcanti et al. 2011) exist, and they may introduce additional complexity to all scenarios discussed above.

The key to solve the above-mentioned problems is treating each citation differently. The goal can be achieved by associating different weights to citations. The higher ranking is, the greater their scholarly impact is. This is the basic principle of using weighted citations to solve the problem of the scholarly impact assessment. Comparing with raw citations, weighted citations are more reasonable for scholarly impact assessment, since it takes account of some latent influential factors of citations and takes full advantage of them.

There have been numerous studies on citation weighting. Many effective solutions to the assessment of the scholarly impact use weighted citations. But there is a lack of a systematic summary of citation weighting methods. The contribution of this paper is to summarize the existing weighting methods that address the assessment of the scholarly impact, as well as, to discover the hidden trends and shortcomings.

The rest of the paper is organized as follows. In "Classic scholarly impact assessment indicators" section, we summarize some of the classic indicators in the field of academic impact assessment. In "Citation weighting solutions for scholarly impact assessment" section, we summarize some of the weighting methods. In "Open issue" section, we present the existing problems and the issues that haven't been taken into consideration. Finally, we provide a conclusion of this paper in "Conclusion" section. The overall structure of this paper is summarized in Fig. 1.

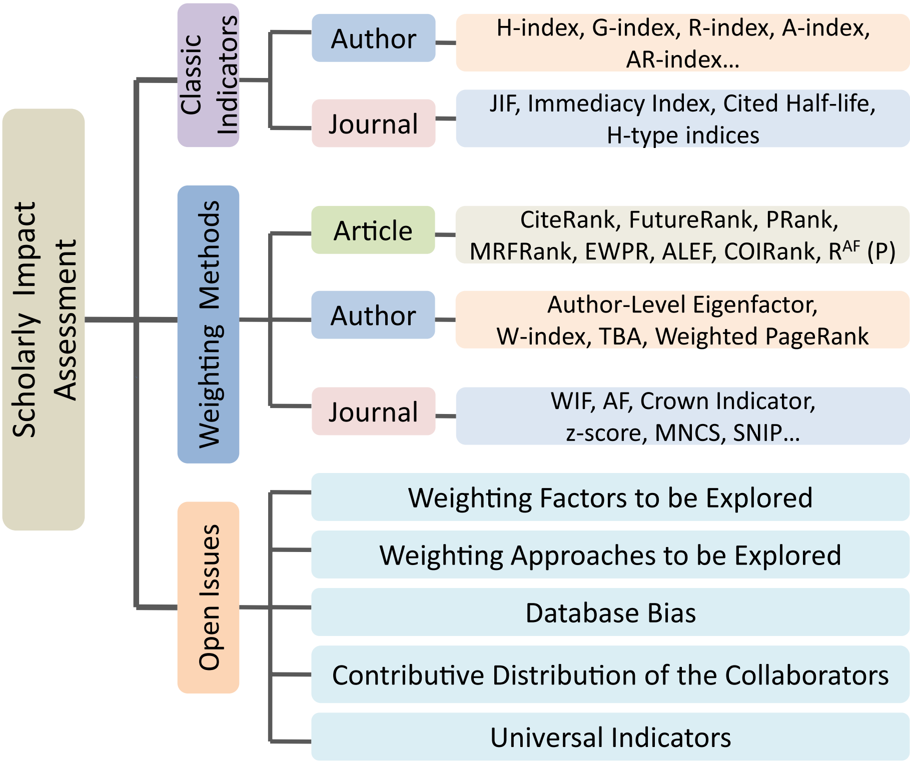

Fig. 1 The structure of scholarly impact assessment

Table 1 Common symbols in the "Classic scholarly impact assessment indicators" section

Symbol	Description
h	The author's H-index
cit_j	The citation number of article j
a_j	The age of article j
$C_{ji}(t, T)$	Number of citations emitted by J_j in t to the articles that J_i publishes in the T
$A_i(t, T)$	Number of articles J_i publishes in T

Classic scholarly impact assessment indicators

In the evaluation of scientific research, we often use some indicators to measure scholarly impact. In this section, we will discuss the classic indicators which are often used in the scholarly impact assessment. We list the commonly used symbols in the following Table 1.

Indicators to evaluate authors

H-index

Among various indicators that assess a scholar's impact, H-index (Hirsch 2005) is one of the most popular indicators for assessing the lifetime achievement of a scholar. It can be calculated as following. A scholar's publications are sorted in the descending order by the number of citations. If there are *h* publications that are cited at least *h* times, the scholar's H-index equals to *h*. H-index is the first indicator for scholar ranking. But its impact actually goes far beyond this. It can also be used to evaluate the impact of journals and institutions. However, it has some obvious shortcomings.

The disadvantages of H-index include: (1) It emphasizes on long term observations, because H-index needs time to accumulate; (2) It can not distinguish the difference among ordinary scientists in the different domain; (3) Because there is no temporal information of citations, it monotonously increase; (4) It is not sensitive to the number of citations received; (5) Apart from this, H-index also suffers from typical problems of citation-based metrics such as: field-dependency, self-citations, identification of scientists, lack of reference standards, and comprehensive data collection, etc (Costas and Bordons 2007; Jin et al. 2007).

G-index

Egghe (2006) proposes G-index which is sensitive to the number of citations received by the first *h* publications of an author. G-index is not only derived from H-index but also an improvement of H-index. It is defined as the highest rank such that the cumulative sum of the number of citations received is larger than or equal to the square of this rank. A higher value of G-index indicates that the scholar has greater scholarly achievement so that his/her impact is much higher. The G-index is a good reflection of the impact of highly cited papers, which overcomes the shortcomings of H-index.

Hg-index

Alonso et al. (2010) propose a new index to characterize the scientific output of researchers called Hg-index, which is calculated based on both H-index and G-index. It fuses all the benefits of both previous measures and minimizes their drawbacks. A researcher's Hg-index is computed as the geometric mean of his/her H-index and G-index. The Hg-index has many advantages. For example, it is not only easy to calculate but also has more granularity than H-index and G-index. In addition, it takes into account the high-cited articles to make up for the shortcomings of the H-index. It can also reduce the impact of single high-cited papers and make up for the shortages of G-index.

A-index, R-index and AR-index

Jin (2006) proposes A-index, which measures the average number of citations for the first *h* publications of an author. A-index is simply defined as the average number of the citations received by the publications which are included in the Hirsch core (Jin et al. 2007). The

A-index has achieved the same goal as G-index. Both of them make up for the shortcomings of H-index that is insensitive to the highly cited papers. Mathematically, A-index can be described as

$$A = \frac{1}{h}\sum_{j=1}^{h} \mathrm{cit}_j \tag{1}$$

where h denotes the author's H-index, cit_j denotes the number of citations in a decreasing order. Then they propose the R-index (Jin et al. 2007), which gives the square root of the cumulated citations in the Hirsch core. Later they propose AR-index (Jin et al. 2007), which regards the age of a publication as a factor. It means that citation counts received by a publication in the h-core are divided by the age of the publication. Mathematically, they can be described as

$$R = \sqrt{\sum\nolimits_{j=1}^{h} \mathrm{cit}_j} \tag{2}$$

$$\mathrm{AR} = \sqrt{\sum\nolimits_{j=1}^{h} \frac{\mathrm{cit}_j}{a_j}} \tag{3}$$

where a_j denotes the age of article j.

Other H-type indices

In order to solve H-index's deficiency and to optimize its performance, a series of other H-type indices are proposed, such as *e-index* (Zhang 2009b), *w-index* (Wu 2010), *h*(2)-*index* (Rousseau 2008), h_α-*index* (Eck and Waltman 2008), *h*(*x*)-*index*, *o-index*, $A+$-*index* (Wang 2013) which all emphasize the value of highly cited papers. The h_T-*index* (Anderson et al. 2008) considers the value of all papers. The h_c-*index* pays more attention to the contribution of new papers cited. The h_I-*index* (Batista et al. 2006) and h_m-*index* (Schreiber 2008) focus on multi-author influence distribution. There are a lot of H-type indices which can remedy some shortcomings of H-index, such as *current-index* (Fiala 2014). We list the focus of some H-type indices in Table 2.

Indicators to evaluate journals

Journal impact factor

Journal Impact Factor (JIF) is an indicator developed by the Institute of Scientific Information (ISI) to quantify journals' impact. It's widely used because of its simplicity. Suppose that there are N journals in total and the i-th journal can be denoted as J_i. Let $C_{ji}(t, T)$ denote the number of citations emitted by J_j in year t to the articles that J_i publishes in the time window $T = \{t-1, \dots, t-T\}$. $A_i(t, T)$ denotes the number of articles J_i publishes in T. Then the J_i's JIF in year t, time window T, can be described as

$$\mathrm{JIF}_i(t, T) = \frac{\sum_{j=1}^{N} C_{ji}(t, T)}{A_i(t, T)} \tag{4}$$

Table 2 The focus of some H-type indices

Index name	Publication time	Highly cited papers	Self-citation	Coauthor's contribution
A		√		
AR		√		
A+		√		
b			√	
ch			√	
Current	√	√		
e		√		
G		√		
hg		√		
h_I				√
h_m				√
h_{ms}			√	
h_s			√	
$h(x)$		√		
$h(2)$		√		
h_α		√		
R	√	√		
o		√		
w		√		
W				√

Therefore, if a journal has a large JIF, it means that the journal publishes more frequently-cited articles, i.e., the journal has a greater impact. Usually, the time window is set as two years. When a longer time window is chosen, JIF will take more previous publications into account so the result will be less immediate (Garfield 2006). For instance, the Journal Citations Reports (JCR) provides 5-year Impact Factor as an alternative.

As JIF becoming widely used, enormous debates are made about it. For one thing, JIF is not only widely used but also misused, producing skewed and misleading results. For example, JIF is misused to assess individual papers, authors, publishers, and institutions. This has caused utilitarianism and even "impact factor mania" (Hall and Page 2015; Casadevall and Fang 2014). For another, JIF could be affected by many latent factors, such as journals' size, type, research field, etc. (Dong et al. 2005; Elliott 2014; Moustafa 2015). Due to these flaws, more and more researchers, even publishers are appealing to deprecate JIF (Casadevall et al. 2016; Callaway 2016).

Immediacy index

Immediacy Index is a journal's average number of citations per article in a year. A higher Immediacy Index indicates that the journal's material is cited and used more quickly (Mathur et al. 2009). So short and rapidly published journals may have an advantage in

Immediacy Index (Ha et al. 2006). Therefore, to be precise, Immediacy Index reflects the timeliness instead of the quality of a journal.

Cited half-life

Cited Half-Life is a useful supplement of JIF, which measures journals' long-term impact (Mathur et al. 2009; Bornmann and Marx 2016). The time length which can be calculated as the number of years for a journal to receive 50% of its current citations. Roldan's experiments (Roldan-Valadez et al. 2018) show that Cited Half-Life provides more accurate assessment in particular fields compared with JIF.

Journal H-index

Inspired by H-index (Hirsch 2005), Braun et al. (2006) propose a H-index for journal, which equals to h if a journal publishes at least h papers receiving h citations. Because journal H-index can reflect the quality of a journal and the quantity of its articles simultaneously, it is more robust and balanced than JIF. Saad's experiment (Saad 2006) shows that journal H-index and JIF are significantly correlated. So they suggest H-index can be an alternative of JIF. Harzing and van der Wal (2009) claim that H-index provides more precise and comprehensive results compared with JIF.

Citation weighting solutions for scholarly impact assessment

So far, many scholars and research groups are committed to the study of scholarly impact using citation weighting. In this section, we will discuss the weighted methods used in the scholarly impact assessment from the perspectives of articles, authors and journals, respectively.

Table 3 Common symbols in the "Citation weighting solutions for article assessment" section

Symbol	Description
x	Actual geographical distances
T_u	The time information on node u
Peak_v	The peak time of node v
C_{ij}	The number of citations from paper i to paper j
w_{ij}	The weight of edge from paper i to paper j
y_i	Paper i's publication year
R_i	Paper i's set of reference papers
$R^{\mathrm{AF}}(p)$	The age- and field-rescaled PageRank
$R^{\mathrm{AF}}(c)$	The age- and field-rescaled citation count
$\mu_i^{\mathrm{AF}}(c)$	The mean of the number of citations to the articles in the same field and of similar age as A_i
$\sigma_i^{\mathrm{AF}}(c)$	The standard deviation of the number of citations to the articles in the same field and of similar age as A_i

Citation weighting solutions for article assessment

Since there are a lot of equations in this section, we list the commonly used symbols in the following Table 3.

Network-based weighting methods

CiteRank The evaluation of the impact of scholarly articles plays a very important role in solving the problems of recruitment decisions, funds distribution and promotion, etc. Walker et al. (2007) propose the CiteRank, which more emphasizes on the new articles. The model points out that random surfers are initially distributed exponentially with age, in favour of more recent publications. So the academic value of these papers should be greater. The CiteRank model is based on the PageRank algorithm. By assigning a higher value to newer publications, this method can make newer publications have a greater probability of being found through random surfing. However, since only citations and time information are used, the method is still unable to obtain a reasonable comparison between recent publications (Wang et al. 2013).

FutureRank We hope to find a new method to rank articles based on predicted future citations, because this will help researchers find high-quality articles easier. Sayyadi and Getoor (2009) propose the FutureRank, a method which is able to combine information about citations, authors and the publication time to effectively rank scientific articles by predicting their future rank of a paper. The model adds a personalized time vector R_{time}, which gives more weight to recently published papers. In the model, the use of authorship provides additional information for the ranking of recent publications. If the author has published many prestigious articles before, then his/her new publication may have a good quality.

P-Rank Ding and Yan (2010) propose a new informetric indicator P-Rank that can be used to measure the status of articles, authors, and journals in the heterogeneous citation network. P-Rank can differentiate the weight of citations. In other words, the P-Rank can differentiate the impact of each citation. In terms of P-Rank, the impact of an article is decided by four factors: the authors who cite this article, the journals that cite this article, the papers that cite this article, and the article's publication time.

MRFRank The content information is very important for measuring the quality of articles, especially for the new articles with only a small number of citations. Generally speaking, the more innovative a paper is, the more novel textual features it has. Wang et al. (2016) first present a burst detection-based method to measure the innovative degree of text feature. They proposed a unified model MRFRank, which ranks the future impact of articles and authors like HITS algorithm. Their biggest contribution is the use of article content information to help identify potentially influential new articles.

Weighted quantum PageRank For objectively measuring the impact of scholarly articles, Bai et al. (2017a) leverage the law of geographic distance and citations between different institutions to weight quantum PageRank algorithm. They suggest that the relationship between citations and the actual geographical distance between institutions is based on the following exponential distribution.

$$z = z_0 + ce^{\frac{-x}{t}} \tag{5}$$

where z is citations, z_0 is the initial value, x is actual geographical distances, c is a constant and t represents time. According to the citation law, a weighted quantum PageRank algorithm is constructed. Finally, they find this weighted quantum PageRank algorithm can better differentiate the impact of scholarly articles compared to PageRank algorithm. Taking the geographical distance between institutions as a weighting factor is a new weighting angle. It inspires us to consider issues from different perspectives.

EWPR Luo et al. (2016) propose an approach called Ensemble enabled Weighted PageRank (EWPR). The specific process of the method can be described as follows. They first proposed Time-Weighted PageRank that extends PageRank by introducing a time decaying factor. The impact weight on the edge is defined as

$$w(u, v) = \begin{cases} 1 & T_u < \mathrm{Peak}_v \\ 1/(ln(e + T_u - \mathrm{Peak}_v))^t & T_u \geq \mathrm{Peak}_v \end{cases} \tag{6}$$

where T_u is the time information on node u, Peak_v is the peak time of node v using the time information of all nodes connecting to v, and t is the decaying factor. Then they use a kind of ensemble methods to assemble the authorities of the heterogeneous entities involved in scholarly articles. And finally they use external data sources to further improve the ranking accuracy. Their study shows that EWPR is a good method for ranking scholarly articles.

ALEF Wesley-Smith et al. (2016) propose Article-Level Eigenfactor (ALEF), a novel citation-based ranking algorithm that can be used to assess the impact of individual articles. At first, they calculate citation-based scores for each paper using ALEF. Next, they use the article scores to generate scores for each author. Then they combine the scores of all the authors of a paper to generate an author score for that paper. Finally, they use the citation and author scores to generate a final score for each paper.

COIRank Although citation-based evaluation approaches have been widely used, identifying the anomalous citation is still a difficult problem. In previous studies, researchers paid less attention to a variety of reasons behind a citation (Wang et al. 2013; Zhou et al. 2007), possibly due to the difficulty of identifying anomalous citation. As we all know the anomalous citation would inevitably cause unfairness and inaccuracy in the process of the article impact assessment. To address this drawback, Bai et al. (2016b) propose a Positive and Negative Conflict of Interest(COI)-based Rank algorithm, named PNCOIRank, to acquire positive COI, negative COI, positive suspected COI, and negative suspected COI relationships. They leverage Conflict of Interest (COI) relationships and team relationships to distinguish different citation weights. Then they propose COIRank (Bai et al. 2016a), a novel method that can not only discover the anomalous citations but also assign a low citing weight to weaken the citation relationship. By comparing with other methods, the COIRank is beneficial to improve the evaluation performance.

Neural network based weighted PageRank Fujimagari and Fujita (2015) propose a novel weighting approach, which can automatically assign weights to edges based on Neural Network. First they divide the citation network into clusters, which can be considered as research fields. Then for each edge in a cluster, four kinds of weighting factors are calculated. Let C_{ij} denote the number of citations from paper i to paper j and w_{ij} denote the weight of edge from paper i to paper j. y_i, R_i, and A_i denote paper i's publication year, set of reference papers and set of authors, respectively. Then the weighting factors can be represented as

1. Citing frequency: $w_{ij} = C_{ij}$
2. Publication years:

$$w_{ij} = \begin{cases} (y_i + y_j)/2 - 1970, & \text{if } (y_i + y_j)/2 < 1970 \\ 0, & \text{else} \end{cases}$$

3. Reference similarity: $w_{ij} = \text{Jaccard}(R_i, R_j) + 1$
4. Author similarity: $w_{ij} = \text{Jaccard}(A_i, A_j) + 1$

Thus, there is a four-dimensional feature for each edge, which is then input into a Neural Network. The weight of each edge is updated according to the output of Neural Network. Then the procedure starts again from clustering until the weight is not updated anymore. Theoretically, this weighting approach can find the optimal combination of the above-mentioned four factors. However, in practice, the model doesn't always find the optimal value.

Age- and field-rescaled PageRank Motivated by the citation z-score and other related works, Vaccario et al. (2017) present the age- and field-rescaled PageRank $R^{\mathrm{AF}}(p)$. Just as its name implies, it takes the field difference and the publication time into account. They also propose the age- and field-rescaled citation count $R^{\mathrm{AF}}(c)$. Specifically, if A_i denotes the i-th article and c_i denotes the number of citations A_i receives, $R^{\mathrm{AF}}(c)$ of A_i is defined as

$$R_i^{\mathrm{AF}}(c) = \frac{c_i - \mu_i^{\mathrm{AF}}(c)}{\sigma_i^{\mathrm{AF}}(c)} \tag{7}$$

where $\mu_i^{\mathrm{AF}}(c)$ and $\sigma_i^{\mathrm{AF}}(c)$ are the mean and standard deviation of the number of citations to the articles in the same field and of similar age as A_i. The articles that have similar age with A_i are defined as Δ_c articles. They are published at the similar time with A_i. They set Δ_c as 1000. $R^{\mathrm{AF}}(p)$ can be calculated by replacing the number of citations in $R^{\mathrm{AF}}(c)$ with A_i's PageRank score. They claim that it performs better than PageRank and age weighted PageRank.

Citation weighting solutions for author assessment

From the author's point of view, the problem of weighted citations can be roughly divided into two major problems. One problem is the self-citation weighting problem, and the other one is the problem of the co-authors' contribution distribution.

Citation-based weighting methods

Self-citations solutions In the scientific community, the citation is usually regarded as a part of the reward system. But self-citations will distort the system. The *ch-index* proposed by Kosmulski (2006), the h_{ms}-*index* and the h_s-*index* proposed by Schreiber (2010) and the *b-index* proposed by Brown (2009), they all remove the number of self-citations in calculating the citation number. It is the equivalent of setting the weight of the self-citations to zero. We will not discuss whether it is reasonable to set the weight of self-citations to zero. One generally recognized point is that self-citations must be distinguished from other references and it should be assigned less weight.

Schubert et al. (2006) propose a fractional method to quantify the weight of self-citation in coauthor relationships. At first, the fractional self-citation counting uses the Jaccard Index to determine the overlapping of co-authors between citations and citing articles. Then it gives more weight to the citation if its citing authors are less overlapped with the cited author.

Cooperative contribution distribution solutions It is very common to have multiple authors in a paper, so how to allocate the contributions of multiple authors is a problem that cannot be ignored. As we all know, H-index is increasingly being used to evaluate the achievements of individual scientists. But when we use the total number of citations or H-index to evaluate the scholarly impact of an author, we default that this author takes full credit for all his papers. But this assumption is obviously unreasonable to papers with multiple authors. Sekercioglu (2008) proposes that the kth ranked co-author is considered to contribute $1/k$ as much as the first author. But there are some flaws in this method. First, he does not carefully consider the contribution of the corresponding author. In fact, the corresponding author plays an important role in the paper, so his/her contribution cannot be ignored. Moreover, a critical flaw in this method is that author weight initially decays quickly, but then becomes almost constant. Whereas ideally they should follow a linear distribution whereby weights are directly proportional to authors' ranking.

Zhang (2009a) proposes W-index that can be used to calculate weighted citation numbers and weighted H-index. The core idea of this scheme is to set the same weight for the first and corresponding authors, and the weights of other authors decrease linearly according to the author's ranking. The weighted citation and the weighted H-index obtained through this weighting scheme can more accurately assess the impact of an author.

Similarly, Galam (2011) proposes Tailor Based Allocations (TBA), by which the first author and the last author (usually the corresponding author) are given a lot of weight and the authors in other locations are also different in weight. Both of the above two methods give the first author and the corresponding author a larger weight, which is more consistent with people's subjective understanding.

Hagen (2010) proposes a method of weighting, Harmonic counting. The higher ranking of the author's name, the greater weight will be given. And the weight of each collaborator decreases with the increase of the number of collaborators. However, this method does not give specific weight to the corresponding author. The solutions to address the contributions of collaborators go far beyond that.

Network-based weighting methods

Weighted PageRank In fact, the citation relationship among papers is a giant net, which connects all articles through citation relationships to form a citation network (Fiala et al. 2008). Since it is not possible for a paper to cite a newly published paper thereafter, this network is a directed acyclic network. Inspired by the PageRank algorithm, many scholars have proposed weighted PageRank algorithms and applied them in the assessment of the author's impact (Nykl et al. 2014, 2015). Yan and Ding (2011) make use of the weighted PageRank algorithm to assess the scholarly impact of the authors from the information retrieval field. They incorporate citation counts with topology of network into the following formula:

$$PR_W(p) = (1-d)\frac{CC(p)}{\sum_{j=1}^{N} CC(p_j)} + d\sum_{i=1}^{k}\frac{PR_W(p_i)}{C(p_i)} \tag{8}$$

where $PR_W(p)$ is the weighted PageRank of author p, $CC(p)$ is the number of citations that the author p received, $\sum_{j=1}^{N} CC(p_j)$ is the all citation counts in the network, d is damping factor, $(1-d)$ is the coefficient to retain the sum of PageRank as one, and $C(p_i)$ is the number of outlinks of p_i. Ding and Yan (2010) extend this algorithm to non-homogeneous citation networks. They propose weighting citations based on factors such as the prestige of the citing journal.

Author-level eigenfactor West et al. (2013) propose using Author-Level Eigenfactor Metrics to assess the impact of authors, institutions and countries within the Social Science Research Network (SSRN) community. As we all know the Eigenfactor score is originally designed for ranking scholarly journals. West et al. (2013) propose it can be adapted to ranking the scholarly output of authors, institutions and countries based on author-level citation data.

TimeRank Franceschet and Colavizza (2018) propose a dynamic approach called TimeRank to rate authors using citations. This method considers the relative position of two authors at the time of the citation among them. It defines a citation reward ρ. The value of ρ is between 0 and 1. If the rating of the citing author is higher than that of the cited author, the citation reward is close to one. Conversely, it tends to be zero. Only when the citing and cited authors have similar ratings, the citation reward is close to 0.5. The ratings of authors at time t are determined by two parts: their previous ratings at time $t-1$ and the citing authors' ratings at time $t-1$. This method takes the dynamics of the citation into account.

Table 4 Common symbols in the Citation weighting solutions for journal assessment" section

Symbol	Description
N	Total number of journals
J_i	The i-th journal
t	Year t
T	Time window $\{t-1, \dots, t-T\}$
$C_{ji}(t,T)$	Number of citations emitted by J_j in t to the articles that J_i publishes in the T
$A_i(t,T)$	Number of articles J_i publishes in T
n	Number of articles in J_i
A_k	The k-th article in J_i
c_k	Number of citations A_k receives

Citation weighting solutions for journal assessment

Since there are a lot of equations in this section, we list the commonly used symbols in the following Table 4.

Citation-based weighting methods

Weighted impact factors A journal's impact is determined by two aspects: counting citations and weighting citations with the prestige of the citing journals. The former can be referred to as *popularity* and the latter as *prestige* (Bollen et al. 2006; Franceschet 2010). Since JIF can only represent a journal's popularity (Bollen et al. 2006), Habibzadeh and Yadollahie (2008) propose Weighted Impact Factor (WIF), which takes journal's relative prestige into account. The definition of WIF is similar to the definition of JIF. The only difference is that the number of citations in JIF is replaced with weighed citations. Specifically, J_i's WIF is defined as

$$\mathrm{WIF}_i(t,T)=\frac{\sum_{j=1}^{N} w_{ji}C_{ji}(t,T)}{A_i(t,T)} \tag{9}$$

where w_{ji} is the weight of $C_{ji}(t,T)$. Apparently, WIF equals to JIF when the weighting factor w_{ji} equals to 1. Moreover, the relative prestige of J_j and J_i is measured by

$$q_{ji}(t,T)=\frac{\mathrm{JIF}_j(t-1,T)}{\mathrm{JIF}_i(t-1,T)} \tag{10}$$

Because the value of $q_{ji}(t,T)$ varies significantly from zero to infinity, the normalized $q_{ji}(t,T)$ is used as the weight. So the weight is defined as

$$w_{ji}=10\cdot\frac{1-0.828\cdot e^{-q_{ji}(t,T)}}{1+16.183\cdot e^{-q_{ji}(t,T)}} \tag{11}$$

Thus, greater weights are assigned to citations from more prestigious journals, i.e., the citations from more prestigious journals are considered to be more important. To get more accurate results, the JIF used to calculate the weight w_{ji} can be replaced by WIF after WIF is calculated for the first time. Habibzadeh and Yadollahie claim that the WIF provides a better "yardstick" compared with JIF. However, Waltman and van Eck (2008) point out that WIF has some serious computational problems, which may cause misleading results.

Later, Zitt and Small (2008) propose Audience Factor (AF). The definition of AF and WIF are basically the same, except that the weight w_{ji}. WIF weights the citation from the perspective of cited-side, which is adequate for the articles in the same field. However, the citing behaviors varies significantly from field to field (Zitt et al. 2005). The source data of Fig. 2 comes from the web of science in 2014. The data presents the number of citations of 27 fields in 2014 and reflects the difference in citing behaviors among different fields. As shown in Fig. 2, Medicine papers have the largest number of citations but its citation per paper is below average, while Multidisciplinary papers are exactly the opposite. So, in contrast to WIF, AF weights the citation from the citing-side, taking journals' citing

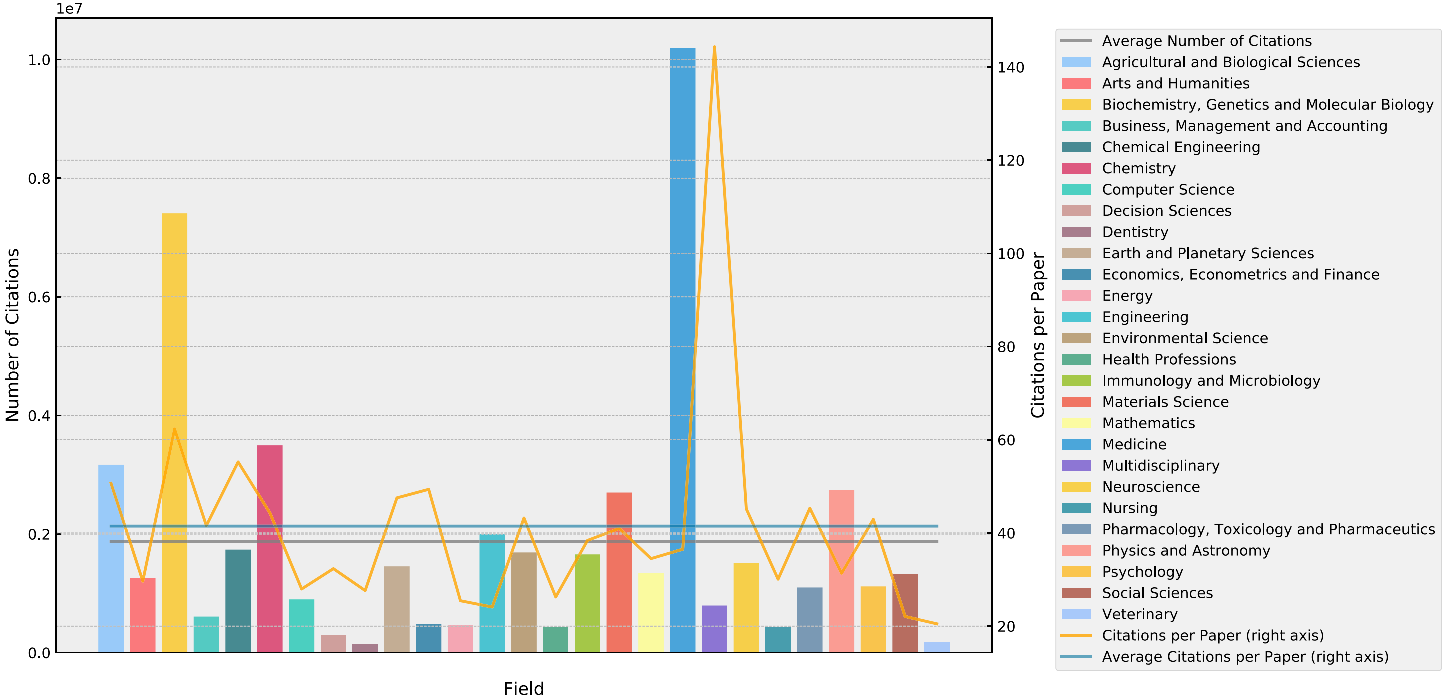

Fig. 2 The difference in citing behaviors among different fields in 2014 (Data from web of science). The units of the left and right axes are different. The unit of the left axis is 10^7

propensity into consideration. The citing propensity is measured by $m_j(t, T)$, the average number of "active" citations emitted by articles in J_j, where active citation is defined as the citations emitted in time window T. Similarly, let $m_s(t, T)$ denote the average number of active citations emitted by articles in all source journals. Then, the weight of $C_{ji}(t, T)$ is calculated as

$$w_{ji} = \frac{m_s(t, T)}{m_j(t, T)} \tag{12}$$

Thus, citations emitted by a journal with a shorter reference list will receive a greater weight. AF is independent from different citing propensities among fields and independent from the field classification system. Furthermore, a field-based variant of AF is also proposed, where the denominator of the weight w_{ji} is replaced with the average number of active citations emitted by articles in a field. Zitt and Small claim that this variant is more robust but it depends on the field classification system.

Leydesdorff and Opthof (2010) introduce another weighting scheme aiming at eliminating field differences. This weighting scheme is referred to as Fractional Citation Counts, which also takes the length of reference list into account. However, instead of weighting citations at journal-level using average reference list length like AF, it weights the citation at the article-level. For example, if an article cites x articles, each of them can be counted as $1/x$ citation, i.e., each citation emitted by that article receives the same weight $1/x$. The citations emitted by articles from different fields will receive different weight due to different citing behaviors. Leydesdorff and Bornmann (2011) claim that JIF calculated based on fractionally counted citations can be interdisciplinary comparable. However, experiments (Waltman and van Eck 2013; Waltman and Van Eck 2013) show that the fractional counting scheme is not properly normalized for different fields.

Crown indicators The Crown Indicator (also known as CPP/FCSm indicator) is a field-weighted indicator created by Centre for Science and Technology Studies (CWTS) to evaluate a group of publications (journals, research groups, etc.). Here, we regard it as an indicator used for journal assessment. The CPP and FCSm stand for citations per publication and the mean field citation score, respectively. Thus the CCP/FCSm indicator of J_i is defined as

$$\mathrm{CPP/FCSm} = \frac{\sum_{k=1}^{n} c_k/n}{\sum_{k=1}^{n} e_k/n} = \frac{\sum_{k=1}^{n} c_k}{\sum_{k=1}^{n} e_k} \tag{13}$$

where e_k denotes the expectation of A_k's number of citations, i.e., the average number of citations to all articles in the same field as A_k. Since the number of citations is normalized by the mean field citation score, CCP/FCSm indicator is interdisciplinary comparable. Lundberg (2007) and Opthof and Leydesdorff (2010) point out that there are some flaws in CCP/FCSm indicator and the most important of which is that it weights the aggregation of articles instead of the individual article. For the journals in the same field, CCP/FCSm indicator is mathematically identical to the sum of citations. Therefore, old publications and reviews still have an advantage in CCP/FCSm indicator. Moed (2010a) argues that this weighting approach is chosen because the overall quality is considered to be more important than the specific distribution of citations.

Lundberg (2007) proposes citation z-score as an alternative to the Crown Indicator, which weights citations in the article-level. It uses z-score to normalize the citations so that the distribution of citations can be considered comprehensively. A_k's normalized number of citations is defined as

$$W_k = \frac{ln(c_k + 1) - \mu}{\sigma} \tag{14}$$

where μ and σ denote the mean value and standard deviation of the natural logarithm of citations to the articles in the same field, respectively. Note that instead of using the number of citations c_k directly, the citation z-score uses the logarithm of the number of citations $ln(c_k)$ to reduce the skewness among fields and journals. However, the use of logarithm may be a drawback for the case that one is only interested in the extreme distribution. Finally, citation z-score is defined as the average normalized citations of all articles in a journal.

Waltman et al. (2011a) present Mean Normalized Citation Score (MNCS) as the new crown indicator, which is another article-level field normalization scheme. It can be considered that CPP/FCSm indicator is the quotient of mean values while MNCS indicator is the mean of quotients. Specifically, the MNCS of J_i is defined as

$$\text{MNCS} = \frac{1}{n}\sum_{k=1}^{n}\frac{c_k}{e_k} \tag{15}$$

Unlike the CCP/FCSm indicator, which gives more weights to the articles from fields with high expected citations, MNCS indicator weights citations more equally. Leydesdorff and Opthof (2011) point out that the field classification system MNCS uses is overlapped and cross-field articles are not well weighted. Nonetheless, experiments made by Smolinsky (2016) prove that MNCS indicator satisfies some basic requirements while keeping computational simplicity.

Waltman et al. (2011b) combine the fundamental principles of MNCS and PageRank-based iterative methods and present a recursive field normalization scheme called Recursive MNCS Indicator. The procedure starts with calculating the non-recursive MNCS value, which is also referred to as the first-order MNCS. Then, in the calculation of second-order MNCS, a greater weight is assigned to the citations from journals with high first-order MNCS. Similarly, the third-order MNCS is calculated based on the second-order MNCS. Although it's sensible to weight citations by field difference and prestige simultaneously, the Recursive MNCS Indicator's performance is not satisfactory. It is too sensitive to the classification system and sub-field characteristics.

SNIP Moed (2010b) introduces a new indicator for journal assessment, called Source Normalized Impact per Paper (SNIP), which can be considered as an extension of AF. So, similar to AF, SNIP is also a citing-side normalization scheme (denoted as "source normalization" in the original article), taking citing propensity (or "citation potential") into account. However, SNIP weights citations in a more comprehensive way, considering the length of reference list, the time required for citations to mature, and the coverage of the database used for assessment. Specifically, SNIP of J_i is defined as

$$\text{SNIP} = \frac{\text{RIP}}{\text{RDCP}} = \frac{\sum_{k=1}^{n} c_k/n}{\text{RDCP}} \tag{16}$$

where RIP and RDCP stand for Raw Impact per Paper and Relative Database Citation Potential, respectively. Database Citation Potential (DCP) of J_i is defined as the average length of reference list of the articles that cite J_i. The DCP of J_i divided by the median DCP value of J_i's subject field, is called RDCP. Note that the subject field of J_i in SNIP scheme is defined as the set of journals that cite J_i. So SNIP is independent from classification system and the differences among specific topic or research scope can also be eliminated.

Leydesdorff and Opthof (2010) point out some deficiencies of the SNIP indicator. They argued that "SNIP involves a transgression of the order of operations in mathematics". Waltman et al. (2013) make a number of modifications to SNIP. The revised SNIP solves the "counterintuitive" problems that the original SNIP has and the basic idea of citation weighting remains the same. Experiments (Pajić 2015; Mingers and Yang 2017) show that the SNIP does best in normalizing field differences.

Network-based weighting methods

Y-factor As mentioned above, JIF can only reflect journals' popularity, rather than prestige. Considering the journal citation relationship can be represented as journal citation network (where nodes denote journals and directed edges denote citations), Bollen et al. (2006) suggest that weighted PageRank algorithm can be used as a indicator of journals' prestige. The basic idea of using PageRank to evaluate journals' prestige is that a journal which is cited by many prestigious journals can be also considered prestigious. By reassigning the prestige through citations iteratively, a stable state will be reached eventually, which can be used to assess journals' prestige. Moreover, Bollen et al. (2006) suggest that the prestige that the citing journal gives out should match to the frequency it cites other journals. Thus, they define the weight of edge from J_i to J_j, i.e., the weighted citations from J_i to J_j as

$$W_{ij} = \frac{C_{ij}}{\sum_{k=1}^{N} C_{ik}} \tag{17}$$

where C_{ij} is the number of citations from J_i to J_j. So, for example, if J_i cites J_j twice and J_k once, J_j will receive 2/3 of J_i's prestige while J_k receive 1/3 of which.

Finally Bollen et al. (2006) propose Y-factor, which is simply defined as the product of a journal's JIF and its weighted PageRank value. Therefore, a journal will have a high Y-factor when it has either high JIF or high weighted PageRank value, or both. Dellavalle et al. (2007) claim Y-factor provides a more refined result than JIF does.

Eigenfactor Bergstrom (2007) proposes Eigenfactor^{TM} metrics to assess the influence of journals, which is also inspired by the PageRank algorithm. Similar to the weighted PageRank part of Y-factor, journals that receive more citations from influential journals are considered more influential. However, Eigenfactor uses field average length of reference list to weight each citation (West et al. 2010). For example, if the citing journal J_i's field cites x articles per paper, the weight $1/x$ is given to each citation J_i emits. Note that in Eigenfactor metrics, self-citations are totally excluded and time window is set as five years. Since a journal's quality is not necessarily related to its size, a size-independent metrics called Article Influence^{TM} Score (AIS) is also proposed. A journal's AIS is defined as its Eigenfactor score divided by the number of articles it publishes.

Experiments made by Davis (2008) and Fersht (2009) show that the Eigenfactor Score is significantly correlated with the total number of citations, i.e., Eigenfactor doesn't provide much more information than raw citation count does. Besides, Waltman and Van Eck (2010) discover a significant correlation between the Eigenfactor and AF. They claim that the two indicators can be regarded as identical for most practical purposes.

SJR indicators The SCImago Journal Rank (SJR) indicator, developed by the SCImago Research Group (Gonzalezpereira et al. 2010), is another indicator inspired by PageRank. Although the way SJR indicator calculates the weight which is the same as the weighed PageRank part of Y-factor, SJR indicator handles the problems of "dangling nodes" (the journals that don't cite other journals) better. The major differences between Eigenfactor and SJR indicator are that they use different databases (Ramin and Shirazi 2012), SJR indicator limits the number of self-citations while Eigenfactor excludes all self-citations, and SJR indicator set time window to three years. Comparisons among JIF, Eigenfactor and SJR indicator are made by different researchers in different fields (Ramin and Shirazi 2012; Kianifar et al. 2014; Cantín et al. 2015). All results suggest that considering all three indicators at the same time is more appropriate.

Later, Guerrero-Bote and Moya-Anegón (2012) suggest SJR2 indicator as an improvement of SJR indicator. In addition to the prestige, the SJR2 indicator also takes the closeness of journals into consideration. The closeness is measured by the cosine between the co-citation profiles of two journals. The larger the cosine is, the closer the themes of the two journals are and greater the weight should be. In other words, the citations from the same field or the same specific subject area are considered to be more important. So, compared with SJR indicator, SJR2 indicator distributes the prestige more equally.

Weighted impact factors In consideration of the structure of citation network, Zitt (2010) develops another variant of AF, which has the advantages of both above-mentioned AF variants. Instead of using articles in a journal or in a field to measure the citing propensity, it uses articles in a specific neighborhood, which is distinguished based on the relation of citations. The neighborhood of J_i is defined as the set of journals cited by J_i. In addition, the second-level neighborhood of J_i is defined as the neighborhood of J_i's neighborhood, including duplicates. Let $x_j(t, T)$ denote the median of J_j's second-level neighborhood reference list length and $x_s(t, T)$ denote the median of reference list length of all source journals' second-level neighborhoods. Then the weight w_{ji} in Eq. 9 is calculated as

$$w_{ji} = k(t, T) \cdot \frac{x_s(t, T)}{x_j(t, T)} \tag{18}$$

where $k(t, T)$ is a correcting factor to equalize the scale of JIF and AF. Since this AF variant can detect research fields based on citation network, it's independent from the classification system, which means it's more robust. However, it still has some limitations. The result is not very accurate for the journals with few references. The citing propensity may be affected by the editorial limitation of the reference list length. Nonetheless, AF can efficiently reduce the bias caused by the different citing propensity among fields.

Zyczkowski (2010) proposes an article-level weighted JIF. This weighting scheme is based on the idea that the citations from a scientist whose articles have already been cited many times are more important than those from a newcomer. Specifically, the weight of each author is calculated based on the author citation network. Suppose there are n_a sample authors in total and the i-th author is denoted as a_i. If the leading eigenvector of the author citation matrix is $x = (x_1, \dots, x_{n_a})$, a_i's weight $w(a_i)$ is calculated as

$$w(a_i) = n_a \frac{x_i}{\sum_{j=1}^{n_a} x_j} \tag{19}$$

Then the weighted citation of A_k is defined as

$$W_k = \sum_{j=1}^{c_k} \frac{1}{n_j} \sum_{\mu=1}^{n_j} w(a_\mu) \tag{20}$$

where n_j is the number of authors of the citing article A_j. Then, the weighted citation of a journal is computed by summing all the weighted citations to the articles it publishes in time window T. This weighting scheme distinguishes the quality of each citation so that the citation inflation can be reduced. However, in practice, it is hard to distinguish identical authors and compute the eigenvector of a huge matrix.

Open issue

Authenticity and fairness are the most basic principles for the scholarly impact assessment. All indicators and methods should allow the principle. On the basis of the existing methods of scholarly impact assessment, there are still some unsolved problems, which can be solved in the future work. In this section, we will give a brief introduction to these issues.

Weighting factors to be explored

We list all weighting factors that are known to us in Table 5. It seems that for different assessing objects (articles, authors or journals), researchers prefer different weighting factors. Journal assessment indicators tend to weight citations by the field or source. However, article assessment indicators often consider the publication time and the citing-side impact, and the author assessment indicators take extra care of self-citations and co-author problems. Vaccario et al. (2017) combine age and field factors for article assessment and the result is positive. So, it's possible to learn what weighting factors can be used for different assessing objects. For example, publication time may be an important weighting factor for journal assessment and the field an author belongs to may have an influence on one's citing behavior. Further research is needed on issues like this.

In addition, there are other factors that are not fully explored but might be useful. Specifically, features such as occurrence frequency (Zhu et al. 2015), time interval, the average length of citing sentences, the average density of citation occurrences (Wan and Liu 2014), the publication language (Ordunamalea and Lopezcozar 2014), the keywords similarity (Fujita et al. 2014), and social network impact (Priem and Hemminger 2010; Costas et al. 2015) might be valuable factors.

Table 5 The factors that can be weighted in the scholarly impact assessment

Evaluated object	Weighting factor
Article	The prestige of the citing author, journal, article (Ding and Yan 2010; Sayyadi and Getoor 2009)
	Publication time, language (Walker et al. 2007) (Sayyadi and Getoor 2009; Fiala 2012; Abujbara et al. 2013)
	Positive, neutral and negative citations (Bai et al. 2016a; Valenzuela et al. 2015; Wan and Liu 2014)
	Geographical distance between institutions (Bai et al. 2017a)
	Belong to the cross field or not (Vaccario et al. 2017)
	Belong to the forefront of research or not (Fujita et al. 2014)
	Location characteristics of citation (Valenzuela et al. 2015)
	The number of citation occurrences (Zhu et al. 2015)
	The average length of citing sentences (Abujbara et al. 2013)
	The average density of citation occurrences (Wan and Liu 2014)
	Reference similarity (Fujimagari and Fujita 2015)
	Author similarity (Fujimagari and Fujita 2015)
Author	The distribution of the co-authors' contribution (Sekercioglu 2008) (Zhang 2009a; Galam 2011; Hagen 2010)
	Self-citation or not (Kosmulski 2006; Schreiber 2010; Brown 2009)
	Authors from the same institution or not (Bai et al. 2016a)
	Authors have worked together or not (Bai et al. 2016a)
Journal	The prestige of the citing journal, author (Habibzadeh and Yadollahie 2008)
	The average length of reference lists in a field or neighborhood (Zitt 2010)
	The number of citations to each article (Leydesdorff and Opthof 2010)
	Sum of average citations of the field each article belongs to (Leydesdorff and Opthof 2010)
	Publication time (Moed 2010b)
	Database coverage (Moed 2010b)
	Closeness of journals (Guerrero-Bote and Moya-Anegón 2012)
	The number of citations that the citing journal emits (Zyczkowski 2010)

Weighting approaches to be explored

Valenzuela et al. (2015) propose a supervised classification approach to identify important citations in scholarly publications. They classify the citations into two types: important citations and incidental citations. They regard the citations which appear in the section of Methods or Discussions as important citations, while the citations of Related Work part as incidental citations. Regardless of the rationality of this classification method, it is a good idea to build an ensemble indicator based on this classification method. To be specific, one can assign weights to citations according to the classification result and then calculate weighted JIF or weighted h-index, etc. Furthermore, Natural Language Processing (NLP) can be used to detect positive, neutral and negative citations or citing purpose (Abujbara et al. 2013; Jha et al. 2017). Similarly, an ensemble assessment indicator can also be built based on it.

Database bias

Many underlying factors could cause database bias. First of all, classification system varies from databases and some of these classification systems are too broad for citation analysis. Thus, for the indicators which depend on classification systems, database bias cannot be ignored. Consequently, different classification systems based on citation relations are proposed (Waltman and van Eck 2012; Gómez-Núñez et al. 2014; Ruiz-Castillo and Waltman 2015). However, there isn't a standard method to evaluate a classification system nor how well an indicator performs on a classification system. Further research is needed to solve this problem.

Secondly, missing the data and the database coverage could affect the assessment results. Whether it uses the number of citations directly or constructing a citation network, one should ensure that the citation data is reliable and authentic so that accurate assessment results can be obtained. Su et al. (2011) present PrestigeRank, using "virtual node" to make up the missing data. PrestigeRank still has limitations and more research is needed.

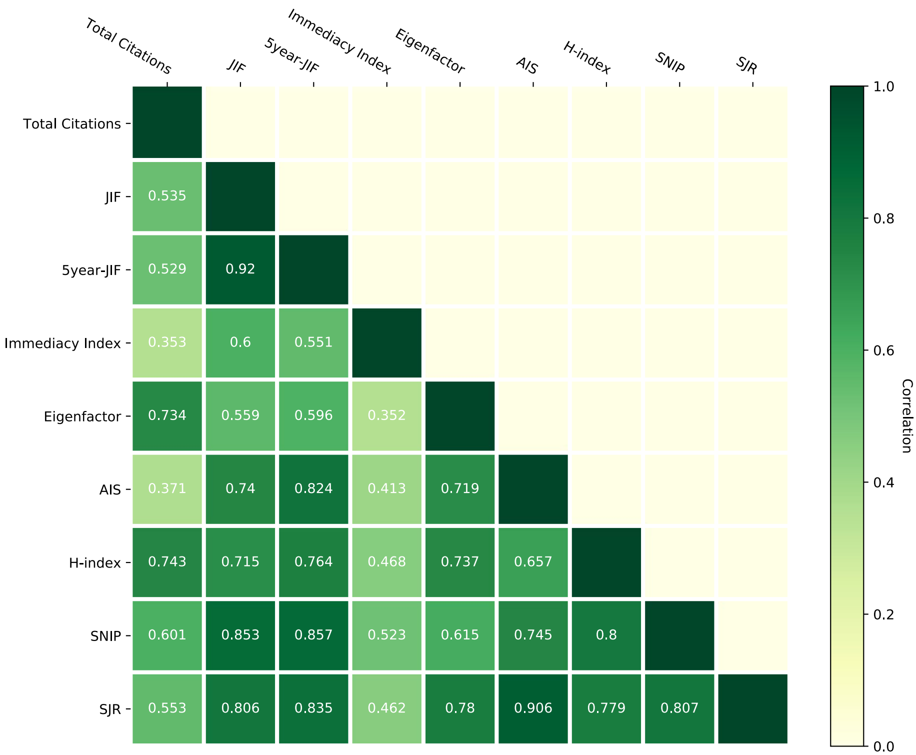

Fig. 3 Correlations between certain indicators, data from Mingers and Yang's experiment in evaluating business and management journals Mingers and Yang (2017)

Contributive distribution of the collaborators

The assignment of academic contributions in collaborative articles is another problem. Although a large number of scholars have put forward methods on this issue, these methods are not necessarily reasonable. A new idea is that when we examine the contribution of the author to an academic paper, we need to consider other related papers published by the author. The greater the influence of the author's other related papers is, the greater the author's contribution to this academic paper is.

Universal indicators

Researchers have been trying to find a perfect indicator for scholarly assessment. Plenty of indicators have been proposed during the past decade. Despite the definitions, weighting factors, and weighting approaches are different, only a few of them provide extra information (Waltman 2016). There are strong correlations between certain indicators, as shown in Fig. 3. Understandably, it's too difficult for a single indicator to cover every aspect. However, perhaps there is a combination of some existing indicators reflecting the scholarly impact from different aspects. It can be used as universal indicator.

Conclusion

Assessment of scholarly impact is very meaningful. It is also one of the most popular trends in academia. There have been countless researchers and research teams who are committed to the study of scholarly impact assessment using key techniques such as statistical analysis, machine learning, data mining or web science. Although new methods and techniques have been constantly emerging, there is still a lot of room for development in the direction of weighted citations. When we use weighted citations to evaluate journals, authors or articles, the results of the evaluation reflect their respective prestige rather than the popularity. Therefore, weighted citations can better reflect scholarly impact of journals, authors, and articles. This review summarizes the existing weighting factors and weighting methods for journals, authors, and articles respectively. After summing up, we find diverse features behind that can cause skewness and unfairness in citation analysis, so a variety of factors are considered to weight citations. Nonetheless, there isn't a perfect weighting approach since it's hard to prove that one indicator outperforms another. In order to continuously improve the performance of indicators and methods to more accurately assess scholarly impact, we can consider the advantages of various methods and the factors that can be weighted. We hope this brief review will be helpful for researchers to find out more reasonable methods of assessing scholarly impact in the future.

Acknowledgements This work was partially supported by the National Natural Science Foundation of China (61872054), the Fund for Promoting the Reform of Higher Education by Using Big Data Technology, Energizing Teachers and Students to Explore the Future (2017A01002), and the Fundamental Research Funds for the Central Universities (DUT18JC09).

References

Abujbara, A., Ezra, J., & Radev, D. R. (2013). Purpose and polarity of citation: Towards NLP-based bibliometrics. In *Proceedings of the 2013 conference of the North American chapter of the association for computational linguistics: Human language technologies* (pp. 596–606).

Alonso, S., Cabrerizo, F. J., Herrera-Viedma, E., & Herrera, F. (2010). hg-index: A new index to characterize the scientific output of researchers based on the h- and g-indices. *Scientometrics*, *82*(2), 391–400.

Anderson, T. R., Hankin, R. K. S., & Killworth, P. D. (2008). Beyond the Durfee square: Enhancing the h-index to score total publication output. *Scientometrics*, *76*(3), 577–588.

Bai, X., Feng, X., Lee, I., Zhang, J., & Ning, Z. (2016a). Identifying anomalous citations for objective evaluation of scholarly article impact. *PLoS ONE*, *11*(9), 1–15.

Bai, X., Hou, J., Du, H., Kong, X., & Xia, F. (2017a). Evaluating the impact of articles with geographical distances between institutions. In *International conference on world wide web companion* (pp. 1243–1244).

Bai, X., Liu, H., Zhang, F., Ning, Z., Kong, X., Lee, I., et al. (2017b). An overview on evaluating and predicting scholarly article impact. *Information*, *8*(3), 73–86.

Bai, X., Zhang, J., Cui, H., Ning, Z., & Xia, F. (2016b). PNCOIRank: Evaluating the impact of scholarly articles with positive and negative citations. In *International conference companion on world wide web* (pp. 9–10).

Batista, P. D., Campiteli, M. G., & Kinouchi, O. (2006). Is it possible to compare researchers with different scientific interests. *Scientometrics*, *68*(1), 179–189.

Bergstrom, C. (2007). Eigenfactor: Measuring the value and prestige of scholarly journals. *College and Research Libraries News*, *68*(5), 314–316.

Bollen, J., Rodriquez, M. A., & Van de Sompel, H. (2006). Journal status. *Scientometrics*, *69*(3), 669–687.

Bollen, J., Sompel, H. V. D., Hagberg, A., & Chute, R. (2009). A principal component analysis of 39 scientific impact measures. *PLoS ONE*, *4*(6), e6022–e6032.

Bornmann, L., & Marx, W. (2016). The journal impact factor and alternative metrics: A variety of bibliometric measures has been developed to supplant the impact factor to better assess the impact of individual research papers. *EMBO Reports*, *17*(8), 1094–1097.

Braun, T., Glänzel, W., & Schubert, A. (2006). A Hirsch-type index for journals. *Scientometrics*, *69*(1), 169–173.

Brown, R. J. C. (2009). A simple method for excluding self-citation from the h-index: The b-index. *Online Information Review*, *33*(6), 1129–1136.

Callaway, E. (2016). Publishing elite turns against impact factor. *Nature*, *535*(7611), 210–211.

Cantín, M., Muñoz, M., & Roa, I. (2015). Comparison between impact factor, eigenfactor score, and scimago journal rank indicator in anatomy and morphology journals. *International Journal of Morphology*, *33*(3), 1183–1188.

Casadevall, A., Bertuzzi, S., Buchmeier, M. J., Davis, R. J., Drake, H., Fang, F. C., et al. (2016). Asm journals eliminate impact factor information from journal websites. *Applied and Environmental Microbiology*, *1*(4), 2407–2408.

Casadevall, A., & Fang, F. C. (2014). Causes for the persistence of impact factor mania. *MBio*, *5*(2), e00,064–14.

Cavalcanti, D. C., Prudncio, R. B. C., Pradhan, S. S., Shah, J. Y., & Pietrobon, R. S. (2011). Good to be bad? Distinguishing between positive and negative citations in scientific impact. In *IEEE international conference on TOOLS with artificial intelligence* (pp. 156–162).

Costas, R., & Bordons, M. O. (2007). The h-index: Advantages, limitations and its relation with other bibliometric indicators at the micro level. *Journal of Informetrics*, *1*(3), 193–203.

Costas, R., Zahedi, Z., & Wouters, P. (2015). Do altmetrics correlate with citations? Extensive comparison of altmetric indicators with citations from a multidisciplinary perspective. *Journal of the Association for Information Science and Technology*, *66*(10), 2003–2019.

Davis, P. M. (2008). Eigenfactor: Does the principle of repeated improvement result in better estimates than raw citation counts? *Journal of the American Society for Information Science and Technology*, *59*(13), 2186–2188.

Dellavalle, R. P., Schilling, L. M., Rodriguez, M. A., van de Sompel, H., & Bollen, J. (2007). Refining dermatology journal impact factors using PageRank. *Journal of the American Academy of Dermatology*, *57*(1), 116–119.

Ding, Y., & Yan, E. (2010). Measuring scholarly impact in heterogeneous networks. *Proceedings of the Association for Information Science and Technology*, *47*(1), 1–7.

Dong, P., Loh, M., & Mondry, A. (2005). The "impact factor" revisited. *Biomedical digital libraries*, *2*(1), 1–8.

Eck, N. J. V., & Waltman, L. (2008). Generalizing the h- and g-indices. *Journal of Informetrics*, *2*(4), 263–271.

Egghe, L. (2006). An improvement of the H-index: The G-index. *ISSI*, *2*(1), 8–9.

Elliott, D. B. (2014). The impact factor: A useful indicator of journal quality or fatally flawed? *Ophthalmic and Physiological Optics*, *34*(1), 4–7.

Fersht, A. R. (2009). The most influential journals: Impact factor and eigenfactor. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(17), 6883–6884.

Fiala, D. (2012). Time-aware pagerank for bibliographic networks. *Journal of Informetrics*, *6*(3), 370–388.

Fiala, D. (2014). Current index: A proposal for a dynamic rating system for researchers. *Journal of the American Society for Information Science and Technology*, *65*(4), 850–855.

Fiala, D., Rousselot, F., & Jezek, K. (2008). Pagerank for bibliographic networks. *Scientometrics*, *76*(1), 135–158.

Franceschet, M. (2010). The difference between popularity and prestige in the sciences and in the social sciences: A bibliometric analysis. *Journal of Informetrics*, *4*(1), 55–63.

Franceschet, M., & Colavizza, G. (2018). Timerank: A dynamic approach to rate scholars using citations. *Journal of Informetrics*, *11*(4), 1128–1141.

Fujimagari, H., & Fujita, K. (2015). Detecting research fronts using neural network model for weighted citation network analysis. *Journal of Information Processing*, *23*(6), 753–758.

Fujita, K., Kajikawa, Y., Mori, J., & Sakata, I. (2014). Detecting research fronts using different types of weighted citation networks. *Journal of Engineering and Technology Management*, *32*, 129–146.

Galam, S. (2011). Tailor based allocations for multiple authorship: A fractional gh-index. *Scientometrics*, *89*(1), 365–379.

Garfield, E. (2006). The history and meaning of the journal impact factor. *JAMA*, *295*(1), 90–93.

Gómez-Núñez, A. J., Batagelj, V., Vargas-Quesada, B., Moya-Anegón, F., & Chinchilla-Rodríguez, Z. (2014). Optimizing SCImago journal & country rank classification by community detection. *Journal of Informetrics*, *8*(2), 369–383.

Gonzalezpereira, B., Guerrerobote, V. P., & De Moyaanegon, F. (2010). A new approach to the metric of journals scientific prestige: The SJR indicator. *Journal of Informetrics*, *4*(3), 379–391.

Guerrero-Bote, V. P., & Moya-Anegón, F. (2012). A further step forward in measuring journals scientific prestige: The SJR2 indicator. *Journal of Informetrics*, *6*(4), 674–688.

Habibzadeh, F., & Yadollahie, M. (2008). Journal weighted impact factor: A proposal. *Journal of Informetrics*, *2*(2), 164–172.

Hagen, N. T. (2010). Harmonic publication and citation counting: Sharing authorship credit equitably not equally, geometrically or arithmetically. *Scientometrics*, *84*(3), 785–793.

Hall, C. M., & Page, S. J. (2015). Following the impact factor: Utilitarianism or academic compliance? *Tourism Management*, *51*, 309–312.

Harzing, A. W., & van der Wal, R. (2009). A google scholar h-index for journals: An alternative metric to measure journal impact in economics and business. *Journal of the American Society for Information Science and Technology*, *60*(1), 41–46.

Ha, T. C., Tan, S. B., & Soo, K. C. (2006). The journal impact factor: Too much of an impact? *Annals-Academy of Medicine Singapore*, *35*(12), 911–916.

Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(46), 16569–16572.

Jha, R., Jbara, A. A., Qazvinian, V., & Radev, D. R. (2017). NLP-driven citation analysis for scientometrics. *Natural Language Engineering*, *23*(1), 93–130.

Jin, B. (2006). H-Index: An evaluation indicator proposed by scientist. *Science Focus*, *1*(1), 8–9.

Jin, B., Liang, L., Rousseau, R., & Egghe, L. (2007). The R-and AR-indices: Complementing the h-index. *Chinese Science Bulletin*, *52*(6), 855–863.

Kianifar, H., Sadeghi, R., & Zarifmahmoudi, L. (2014). Comparison between impact factor, eigenfactor metrics, and scimago journal rank indicator of pediatric neurology journals. *Acta Informatica Medica*, *22*(2), 103–106.

Kong, X., Mao, M., Wang, W., Liu, J., & Xu, B. (2018). Voprec: Vector representation learning of papers with text information and structural identity for recommendation. *IEEE Transactions on Emerging Topics in Computing*, *PP*(99), 1.

Kosmulski, M. (2006). A new Hirsch-type index saves time and works equally well as the original h-index. *ISSI Newsletter*, *2*(3), 4–6.

Leydesdorff, L., & Bornmann, L. (2011). How fractional counting of citations affects the impact factor: Normalization in terms of differences in citation potentials among fields of science. *Journal of the American Society for Information Science and Technology*, *62*(2), 217–229.

Leydesdorff, L., & Opthof, T. (2010). Normalization at the field level: Fractional counting of citations. *Journal of Informetrics*, *4*(4), 644–646.

Leydesdorff, L., & Opthof, T. (2011). Remaining problems with the "New Crown Indicator" (MNCS) of the CWTS. *Journal of Informetrics*, *5*(1), 224–225.

Lundberg, J. (2007). Lifting the crown—citation z-score. *Journal of Informetrics*, *1*(2), 145–154.

Luo, D., Gong, C., Hu, R., Duan, L., & Ma, S. (2016). Ensemble enabled weighted PageRank. arXiv preprint arXiv:160405462.

Mathur, V. P., Sharma, A., et al. (2009). Impact factor and other standardized measures of journal citation: A perspective. *Indian Journal of Dental Research*, *20*(1), 81–85.

Mingers, J., & Yang, L. (2017). Evaluating journal quality: A review of journal citation indicators and ranking in business and management. *European Journal of Operational Research*, *257*(1), 323–337.

Moed, H. F. (2010a). CWTS crown indicator measures citation impact of a research group's publication oeuvre. *Journal of Informetrics*, *4*(3), 436–438.

Moed, H. F. (2010b). Measuring contextual citation impact of scientific journals. *Journal of Informetrics*, *4*(3), 265–277.

Moustafa, K. (2015). The disaster of the impact factor. *Science and Engineering Ethics*, *21*(1), 139–142.

Nykl, M., Campr, M., & Jezek, K. (2015). Author ranking based on personalized pagerank. *Journal of Informetrics*, *9*(4), 777–799.

Nykl, M., Jezek, K., Fiala, D., & Dostal, M. (2014). Pagerank variants in the evaluation of citation networks. *Journal of Informetrics*, *8*(3), 683–692.

Opthof, T., & Leydesdorff, L. (2010). Caveats for the journal and field normalizations in the CWTS ("Leiden") evaluations of research performance. *Journal of Informetrics*, *4*(3), 423–430.

Ordunamalea, E., & Lopezcozar, E. D. (2014). Google scholar metrics evolution: An analysis according to languages. *Scientometrics*, *98*(3), 2353–2367.

Pajić, D. (2015). On the stability of citation-based journal rankings. *Journal of Informetrics*, *9*(4), 990–1006.

Priem, J., & Hemminger, B. H. (2010). Scientometrics 2.0: New metrics of scholarly impact on the social web. *First Monday*, *15*(7), 1.

Ramin, S., & Shirazi, A. S. (2012). Comparison between Impact factor, SCImago journal rank indicator and eigenfactor score of nuclear medicine journals. *Nuclear Medicine Review*, *15*(2), 132–136.

Roldan-Valadez, E., Orbe-Arteaga, U., & Rios, C. (2018). Eigenfactor score and alternative bibliometrics surpass the impact factor in a 2-years ahead annual-citation calculation: A linear mixed design model analysis of radiology, nuclear medicine and medical imaging journals. *La Radiologia Medica*, *123*, 1–11.

Rousseau, R. (2008). Woeginger's axiomatisation of the h-index and its relation to the g-index, the h(2)-index and the R2-index. *Journal of Informetrics*, *2*(4), 335–340.

Ruiz-Castillo, J., & Waltman, L. (2015). Field-normalized citation impact indicators using algorithmically constructed classification systems of science. *Journal of Informetrics*, *9*(1), 102–117.

Saad, G. (2006). Exploring the h-index at the author and journal levels using bibliometric data of productive consumer scholars and business-related journals respectively. *Scientometrics*, *69*(1), 117–120.

Sayyadi, H., & Getoor, L. (2009). Futurerank: Ranking scientific articles by predicting their future PageRank. In *Siam international conference on data mining* (pp. 533–544).

Schreiber, M. (2008). EDITORIAL: To share the fame in a fair way, hm modifies h for multi-authored manuscripts. *New Journal of Physics*, *10*(4), 1131–1137.

Schreiber, M. (2010). The influence of self-citation corrections and the fractionalised counting of multi-authored manuscripts on the Hirsch index. *Annalen Der Physik*, *18*(9), 607–621.

Schubert, A., Glänzel, W., & Thijs, B. (2006). The weight of author self-citations. A fractional approach to self-citation counting. *Scientometrics*, *67*(3), 503–514.

Sekercioglu, C. H. (2008). Quantifying coauthor contributions. *Science*, *322*(5900), 371–371.

Smolinsky, L. (2016). Expected number of citations and the crown indicator. *Journal of Informetrics*, *10*(1), 43–47.

Su, C., Pan, Y. T., Zhen, Y. N., Ma, Z., Yuan, J. P., Guo, H., et al. (2011). PrestigeRank: A new evaluation method for papers and journals. *Journal of Informetrics*, *5*(1), 1–13.

Vaccario, G., Medo, M., Wider, N., & Mariani, M. S. (2017). Quantifying and suppressing ranking bias in a large citation network. *Journal of Informetrics*, *11*(3), 766–782.

Valenzuela, M., Ha, V., & Etzioni, O. (2015). Identifying meaningful citations. In *Association for the advancement of artificial intelligence* WS-15-13 (pp. 21–26).

Van, H. B. A., Phelps, J., Barnes, M., & Suk, W. A. (2000). Evaluating scientific impact. *Environmental Health Perspectives*, *108*(9), 392–393.

Walker, D., Xie, H., Yan, K. K., & Maslov, S. (2007). Ranking scientific publications using a model of network traffic. *Journal of Statistical Mechanics: Theory and Experiment*, *6*(6), 1–5.

Waltman, L. (2016). A review of the literature on citation impact indicators. *Journal of Informetrics*, *10*(2), 365–391.
Waltman, L., & van Eck, N. J. (2008). Some comments on the journal weighted impact factor proposed by Habibzadeh and Yadollahie. *Journal of Informetrics*, *2*(4), 369–372.
Waltman, L., & Van Eck, N. J. (2010). The relation between eigenfactor, audience factor, and influence weight. *Journal of the Association for Information Science and Technology*, *61*(7), 1476–1486.
Waltman, L., & van Eck, N. J. (2012). A new methodology for constructing a publication-level classification system of science. *Journal of the American Society for Information Science and Technology*, *63*(12), 2378–2392.
Waltman, L., & van Eck, N. J. (2013a). Source normalized indicators of citation impact: An overview of different approaches and an empirical comparison. *Scientometrics*, *96*(3), 699–716.
Waltman, L., & Van Eck, N. J. (2013b). A systematic empirical comparison of different approaches for normalizing citation impact indicators. *Journal of Informetrics*, *7*(4), 833–849.
Waltman, L., Van Eck, N. J., Van Leeuwen, T. N., & Visser, M. S. (2013). Some modifications to the SNIP journal impact indicator. *Journal of Informetrics*, *7*(2), 272–285.
Waltman, L., Van Eck, N. J., Van Leeuwen, T. N., Visser, M. S., & Van Raan, A. F. J. (2011a). Towards a new crown indicator: Some theoretical considerations. *Journal of Informetrics*, *5*(1), 37–47.
Waltman, L., Yan, E., & van Eck, N. J. (2011b). A recursive field-normalized bibliometric performance indicator: An application to the field of library and information science. *Scientometrics*, *89*(1), 301–314.
Wang, l. (2013). A new h-type index A+ index. *Information Magazine*, *1*, 55–58.
Wang, Y., Tong, Y., & Zeng, M. (2013). Ranking scientific articles by exploiting citations, authors, journals, and time information. In *AAAI conference on artificial intelligence* (pp. 933–939).
Wang, S., Xie, S., Zhang, X., Li, Z., Yu, P. S., & Shu, X. (2016). Future influence ranking of scientific literature. *Computer Science*, *9*, 1–9.
Wan, X., & Liu, F. (2014). Are all literature citations equally important? Automatic citation strength estimation and its applications. *Journal of the Association for Information Science & Technology*, *65*(9), 1929–1938.
Wesley-Smith, I., Bergstrom, C. T., & West, J. D. (2016). Static ranking of scholarly papers using article-level eigenfactor (ALEF). arXiv preprint arXiv:160608534.
West, J. D., Bergstrom, T. C., & Bergstrom, C. T. (2010). The eigenfactor metrics™: A network approach to assessing scholarly journals. *College and Research Libraries*, *71*(3), 236–244.
West, J. D., Jensen, M. C., Dandrea, R. J., Gordon, G. J., & Bergstrom, C. T. (2013). Author-level eigenfactor metrics: Evaluating the influence of authors, institutions, and countries within the social science research network community. *Journal of the Association for Information Science and Technology*, *64*(4), 787–801.
Wu, Q. (2010). The w-index: A measure to assess scientific impact by focusing on widely cited papers. *Journal of the American Society for Information Science and Technology*, *61*(3), 609–614.
Yan, E., & Ding, Y. (2011). Discovering author impact: A PageRank perspective. *Information Processing and Management*, *47*(1), 125–134.
Zhang, C. T. (2009a). A proposal for calculating weighted citations based on author rank. *Embo Reports*, *10*(5), 416–417.
Zhang, C. T. (2009b). The e-index, complementing the h-index for excess citations. *PLoS ONE*, *4*(5), e5429–e5432.
Zhou, D., Orshanskiy, S. A., Zha, H., & Giles, C. L. (2007). Co-ranking authors and documents in a heterogeneous network. In *IEEE international conference on data mining* (pp. 739–744).
Zhu, X., Turney, P., Lemire, D., & Vellino, A. (2015). Measuring academic influence: Not all citations are equal. *Journal of the Association for Information Science and Technology*, *66*(2), 408–427.
Zitt, M. (2010). Citing-side normalization of journal impact: A robust variant of the Audience Factor. *Journal of Informetrics*, *4*(3), 392–406.
Zitt, M., Ramanana-Rahary, S., & Bassecoulard, E. (2005). Relativity of citation performance and excellence measures: From cross-field to cross-scale effects of field-normalisation. *Scientometrics*, *63*(2), 373–401.
Zitt, M., & Small, H. (2008). Modifying the journal impact factor by fractional citation weighting: The audience factor. *Journal of the American Society for Information Science and Technology*, *59*(11), 1856–1860.
Zyczkowski, K. (2010). Citation graph, weighted impact factors and performance indices. *Scientometrics*, *85*(1), 301–315.